

## Utilização de Ferramentas de Bioinformática na Construção de Primers para Detecção de Sequências Específicas de DNA



```
CCAAGAAGGTTAAAGATGCAGTCAAAAGATTGAGGACTAATTGCATCAAGAACAC
AGA GAAAGA CATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAA
GGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAG
TTCCTACTGAATCTAAGGCCATGCATGGAGTCTAAGATTCAAATCGAGGATCTAAC
AGAACTCGCGTGAAGACTGGCGAACAGTTCATACAGAGTCTTTACGACTCAAT
GACAAGAAGAAAAATCTTCGTCAAGATGCTGGAGCACGACACTCTGGTCTACTCCA
AAAAATGTCAAAGATACAGTCTCAAGAAGACCAAAGGGCTATTGAGACTTTCAACA
AAGGAIAAIIICGGGAAACCIICICGGAIICCAIIGCCGAGCIAICIGICACIIICAI
CGAAAGGACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAA
AGGAAAGGCTATCATTCAAGATCTCTCTGCCGACAGTGGTCCCAAAGATGGACCC
CCACCCACGA GGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCITCAAAGC
AARTGGATTGATGTGACATCTCCACTGACGTAAGGATGACGACACAATCCCACTA
TCCTTCGCAAGACCCCTTCTCTATATcAAGCTATGGCGCAGCGCTCCGGTA GAGAG
CTCAG
```

*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Agroindústria de Alimentos  
Ministério da Agricultura, Pecuária e Abastecimento*

## ***Documentos 114***

# **Utilização de Ferramentas de Bioinformática na Construção de Primers para Detecção de Sequências Específicas de DNA**

*Edna Maria Moraes Oliveira  
Tatiane Corrêa de Oliveira  
Ivanilda Santos de Lima  
Thiago Ferreira dos Santos*

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Agroindústria de Alimentos**

Av. das Américas, 29.501 - Guaratiba

CEP: 23020-470 - Rio de Janeiro - RJ

Telefone: (21) 3622-9600

Fax: (21) 2410-1090 / 3622-9713

Home Page: [www.ctaa.embrapa.br](http://www.ctaa.embrapa.br)

E-mail: [sac@ctaa.embrapa.br](mailto:sac@ctaa.embrapa.br)

**Comitê Local de Publicações e Editoração da Unidade**

Presidente: Virgínia Martins da Matta

Membros: Andre Luis do Nascimento Gomes, Daniela De Grandi Castro Freitas,  
Ilana Felberg, Luciana Sampaio de Araújo, Marília Penteado Stephan,  
Michele Belas Coutinho, Renata Galhardo Borguini, Renata Torrezan

Supervisão editorial: Virgínia Martins da Matta

Revisão de texto: Edmar das Mercês Penha

Normalização bibliográfica: Luciana Sampaio de Araújo

Tratamento de ilustrações: Marcos Moulin e Andre Luis do Nascimento Gomes

Editoração eletrônica: Marcos Moulin, Andre Luis do Nascimento Gomes e

Chris Maciel

Ilustração da capa: Chris Maciel

**1ª edição**

1ª impressão (2011): 200 exemplares

**Todos os direitos reservados.**

A reprodução não-autorizada desta publicação, no todo ou em parte,  
constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**

**Embrapa Agroindústria de Alimentos**

---

Oliveira, Edna Maria Morais.

Utilização de ferramentas de bioinformática na construção de primers para  
detecção de sequências específicas de DNA / Edna Maria Morais Oliveira... [et  
al.]. – Rio de Janeiro : Embrapa Agroindústria de Alimentos, 2011.

20p. ; 21 cm. – (Documentos / Embrapa Agroindústria de Alimentos, ISSN  
1516-8247 ; 114).

1. Oligonucleotídeo. 2. DNA. 3. Bioinformática. I. Oliveira, Edna Maria Morais.  
II. Oliveira, Tatiane Corrêa de. III. Lima, Ivanilda Santos de. IV. Ferreira, Thiago.  
V. Série.

CDD 572.8 (21. ed.)

---

©Embrapa 2011

# Autores

## **Edna Maria Morais Oliveira**

Engenheira Química, D.Sc. em Bioquímica, pesquisadora da Embrapa Agroindústria de Alimentos, Rio de Janeiro, RJ, edna@ctaa.embrapa.br

## **Tatiane Corrêa de Oliveira**

Técnica em Alimentos, Assistente da Embrapa Agroindústria de Alimentos, Rio de Janeiro, RJ, correa@ctaa.embrapa.br

## **Ivanilda Santos de Lima**

Nutricionista, Bolsista DTI-III CNPq, Rio de Janeiro, RJ, ivanildalima@gmail.com

## **Thiago Ferreira dos Santos**

Biólogo, Mestrando em Ciência de Alimentos na Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, thfsctaa@gmail.com



# **Apresentação**

O desenho de oligonucleotídeos iniciadores (primers) é uma etapa fundamental para o sucesso de métodos baseados na reação em cadeia da DNA polimerase (PCR).

A demanda por métodos mais sensíveis, específicos e rápidos, para avaliar a qualidade e a segurança de produtos alimentícios, promove o desenvolvimento de métodos moleculares, em especial os baseados em PCR. O uso de programas de Bioinformática com “livre acesso” permite uma seleção mais criteriosa dos primers para a detecção de sequências específicas de DNA.

A partir da experiência adquirida no Laboratório de Diagnóstico Molecular e Micologia da Embrapa Agroindústria de Alimentos, foi organizado o presente Documento cuja principal utilidade é para os usuários do próprio laboratório, mas que, certamente, servirá de orientação para outros laboratórios que se ocupem da detecção de sequências específicas de DNA de diferentes organismos.

Se assim for, nossos esforços voltados para a disponibilização da informação e transferência de conhecimento terão, mais uma vez, valido à pena.

*Regina Celi Araujo Lago*  
Chefe Geral  
Embrapa Agroindústria de Alimentos



# Sumário

<b>Introdução .....</b>	<b>9</b>
<b>Desenvolvimento.....</b>	<b>10</b>
NCBI.....	10
BLAST .....	13
Clustal W .....	13
Repeat Masker .....	15
Gene Fisher .....	15
Bioinfo.net .....	17
<b>Considerações Finais .....</b>	<b>18</b>
<b>Referências .....</b>	<b>18</b>
<b>Literatura Recomendada.....</b>	<b>19</b>





# Utilização de Ferramentas de Bioinformática na Construção de Primers para Detecção de Sequências Específicas de DNA

*Edna Maria Morais Oliveira*

*Tatiane Corrêa de Oliveira*

*Ivanilda Santos de Lima*

*Thiago Ferreira dos Santos*

---

## Introdução

O Laboratório de Diagnóstico Molecular e Micologia da Embrapa Agroindústria de Alimentos vem usando programas de “livre acesso” para buscar sequências de DNA de diferentes espécies e, posteriormente, tratar tais sequências para a definição de oligonucleotídeos iniciadores (primers).

A especificidade de um oligonucleotídeo iniciador é fundamental para o sucesso da Reação em Cadeia da DNA Polimerase (PCR) e, consequentemente, de metodologias baseadas nesta reação.

Sendo assim, a bioinformática é imprescindível para a manipulação de dados biológicos e pode ser definida como uma ferramenta que abrange todos os aspectos de aquisição, processamento, armazenamento, distribuição, análise, e interpretação da informação biológica. Através da combinação de procedimentos e técnicas de matemática, estatística e ciência da computação são elaboradas várias ferramentas que nos auxiliam a compreender o significado biológico dos dados genômicos. Além disso, através da criação de banco de dados com as informações já processadas, acelera-se a investigação em áreas como medicina, biotecnologia, agronomia etc. (QU et al., 2009).

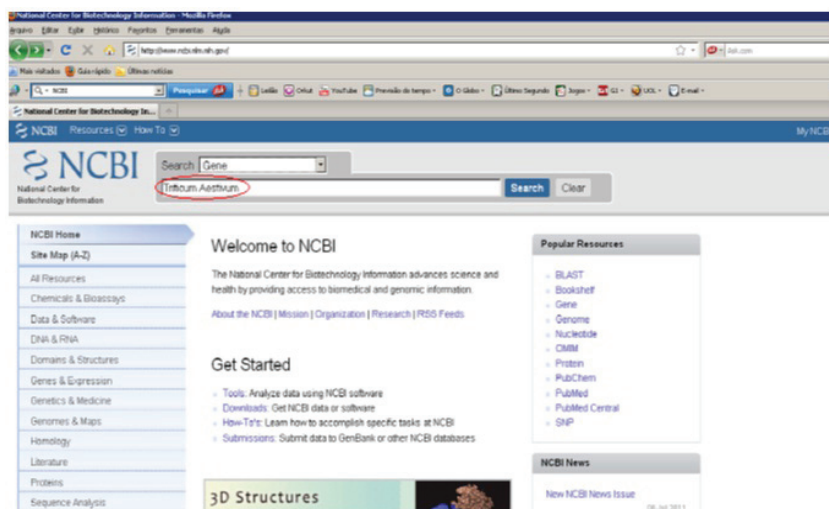
## Desenvolvimento

Devido ao grande número de dados genômicos gerados em laboratórios de todo o mundo, foi necessário organizá-los de maneira acessível de modo a evitar redundância na pesquisa científica e possibilitar a análise por um maior número de cientistas. A construção de bancos de dados para armazenamento de informações de sequências de DNA e genomas inteiros, proteínas e suas estruturas tridimensionais, bem como de vários outros produtos da era genômica é extremamente importante e, ao mesmo tempo, um grande desafio.

A seguir, são apresentados os principais bancos de dados e programas de livre acesso na internet que permitem uma busca específica de sequências de DNA.

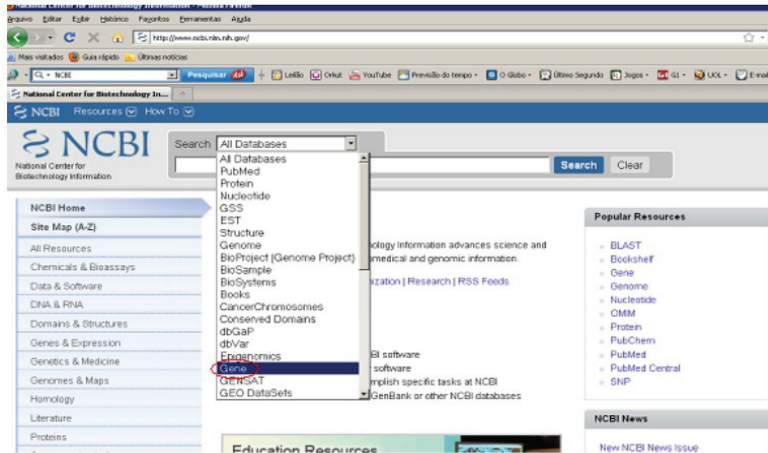
**NCBI** (<http://www.ncbi.nlm.nih.gov/>)

O National Center for Biotechnology Information (NCBI) dos EUA é um banco de dados central sobre informações genômicas, onde se realiza a busca por sequências de interesse. Vários outros bancos de dados similares estão distribuídos por países da Europa e no Japão, que trocam dados a cada 24 horas com o NCBI. O GenBank é o principal banco de dados do NCBI e armazena todas as sequências disponíveis publicamente de DNA, RNA e proteínas.

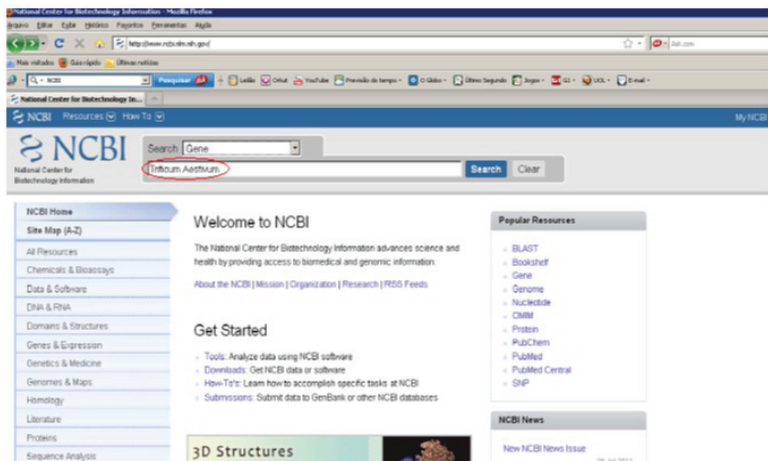


O processo de busca é realizado de acordo com a sequência abaixo.

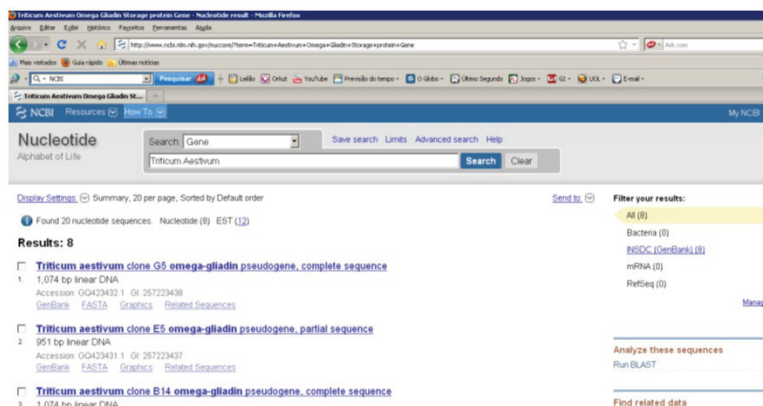
**Passo 1:** Acessar o site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) e selecionar o banco de dados.



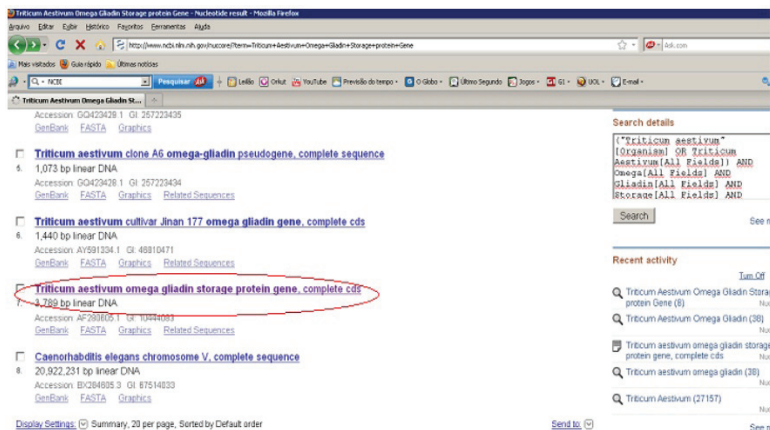
**Passo 2:** Inserir o nome em latim do organismo que se pretende estudar e clicar em **SEARCH**.



**Passo 3:** O banco apresentará uma série de opções para a escolha do organismo a ser estudado.



**Passo 4:** Escolha uma opção entre as apresentadas.



**Passo 5:** Clicar em FASTA, pois a sequência neste formato será usada para as etapas seguintes.



**BLAST** (<http://www.blast.ncbi.nlm.nih.gov/>)

O Basic Local Alignment Search Tool é a ferramenta mais popular de comparação de sequências de DNA com os bancos de dados genômicos. Através deste algoritmo, pode-se comparar uma sequência de DNA ou proteína qualquer com todas as sequências genômicas de domínio público.

É importante salientar que o programa não realiza uma comparação da extensão total das moléculas avaliadas ou consideradas, mas apenas identifica, no banco de dados, a presença de uma sequência suficientemente parecida com a pesquisada, descartando rapidamente os resultados não produtivos, além da extensão à vizinhança da região de homologia, até que isso não seja mais possível (SANTOS; ORTEGA, 2003).

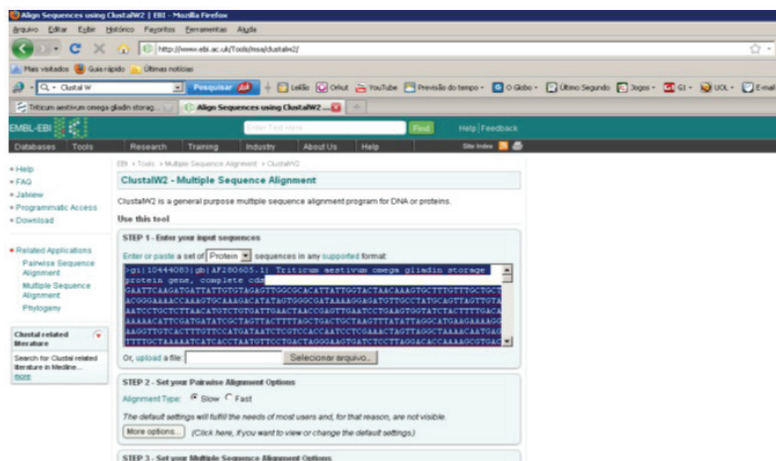
**Clustal W** (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)

É um programa de alinhamento múltiplo de sequências de DNA, RNA e proteínas, onde estes se organizam para comparação e determinação de seu grau de similaridade de forma mais detalhada, sequências muitas vezes divergentes em significância biológica.

O programa calcula os melhores pareamentos para cada sequência e mostra com clareza, nos alinhamentos, as similaridades e diferenças,

como evidenciado no procedimento apresentado em seguida. As relações evolutivas podem ser vistas por cladogramas e filogramas gerados pelo próprio programa.

Acessar o site <http://www.ebi.ac.uk/Tools/msa/clustalw2/> para submissão da sequência que se deseja estudar.



*O resultado se dá da seguinte forma:*

```

rye 1      GGCTAAGCGGCTGGTCTCTTTGCGGCA  Região Homóloga  GCTC  191
rye 2      GGCTAAGCAGCTGGTCTCTTTGCGGCGAGTAGTTGTCGC  CCTCGTGGCTC  191
spelt      GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTTGTCGC  CCTCGTGGCTC  191
rye 3      GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTTGTCGC  CCTCGTGGCTC  191
wheat 1    GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTTGTCGC  CCTCGTGGCTC  191
kamut      GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTCGTCGC  CCTCGTGGCTC  179
wheat 2    GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTCGTCGC  CCTCATGGCTC  596
barley     GGCTAAGCGGCTGGTCTCTTTGCGGCGAGTAGTCGTCGC  CCTCGTGGCTC  486

          → Região Homóloga
rye 1      TCACCGTCGCTSAAGGTGAGGCC-----  214
rye 2      TCACCGTCGCTSAAGGTGAGGCC-----  214
spelt      TCACCGTCGCTSAAGGTGAGGCC-----  214
rye 3      TCACCGTCGCTSAAGGTGAGGCC-----  214
wheat 1    TCACCGTCGCTSAAGGTGAGGCC-----  214
kamut      TCACCGCGCTSAAGGTGAGGCC-----  202
wheat 2    TCACCGCGCTSAAGGTGAGGCC-----  619
barley     TCACCGCGCTSAAGGTGAGGCC-----  536

rye 1      ---TCTGGTCAACTACAGTGTGAGCGCGAGCTCCAGGAGCGTGAGCTCGA  261
rye 2      ---TCTGGTCAACTACAGTGTGAGCGCGAGCTCCAGGAGCGTGAGCTCGA  261
spelt      ---TCTGGTCAACTACAGTGTGAGCGCGAGCTCCAGGAGCGTGAGCTCGA  261
rye 3      ---TCTGGTCAACTACAGTGTGAGTGCAGCTCCAGGAGCGTGAGCTCGA  261
wheat 1    ---TCTGGTCAACTACAGTGTGAGCGCGAGCTCCAGGAGCGTGAGCTCGA  261
kamut      ---TCTGGACAACTACAATGTGAGCGCGAGCTCCGGGAAGCGCGAGCTCGA  249
wheat 2    ---TCTGGACAACTACAATGTGAGCGCGAGCTCCGGGAAGCGCGAGCTCGA  666
barley     CGCTCTAGGCAGCTACAGTGTGAGCGCGAGCTCCAGGAGAGCTCGCTCGA  586
  
```

## Repeat Masker (<http://www.repeatmasker.org/>)

É um programa utilizado para mascarar regiões com sequências de bases nitrogenadas muito repetidas e regiões de baixa complexidade.

## Gene Fisher (<http://bibiserv.techfak.uni-bielefeld.de/genefisher2/submission.html>)

Sendo a sequência de nucleotídeos conhecida e estabelecida, o programa é capaz de desenhar e disponibilizar oito opções de primers para amplificar fragmentos ou um gene inteiro pela técnica de PCR. O programa oferece opções para definir as características dos primers, incluindo o tamanho do produto, tamanho e Tms dos próprios primers, conteúdo de citosina e guanina e a escolha dos primers internos.

**BiBiServ**  
Bielefeld University Bioinformatics Server

Tools Education Administration

**Tools**  
Genome Comparison  
Check  
REPlot  
more  
Alignments  
RefSeqSearch2  
Oxam  
more  
Primer Design  
GeneFisher2  
RNA Studio  
RNA-align  
KnotInFrame  
DNHybrid  
more  
Evolutionary Relationship  
ROSE  
more  
Others  
Kot DB  
JFEditor  
more

**GeneFisher2 - Submission**

Enter the sequences of your project below.

Your input can be a single or multiple sequence(s), either nucleotide or amino acid. For multiple sequences a multiple alignment will be calculated. You can also upload already aligned sequences. Note that sequences are only accepted in [FASTA](#) format.

Click here for [example DNA](#) sequences.

sequence upload:  
[EcoRI](#) [Nanhum a. clonado](#)

paste files in here:

```
>gi12444881:gb|AF280685.1| Tricinum aestivum omega gliadin storage
protein gene, complete cds
GAAATCAGAGTATATTTGTTGAGATTGAGGACATTAATGTTACTACAAAGTCTTTGTTCTCT
ACGGGAAACCAAGTSCAAGGACATATGAGGACATTAAGGAGATGTTCTCTATGCTTTGTTCT
AATCTCTCTCTGATGATGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
AAAAACATTCGATGATGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
AAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
```

[check input](#)  
[Reset](#)

Note: You more than 100 sequences are allowed

Welcome  
Submission  
References  
Manual  
Contact

Primeiro passo se dá por meio da inserção da sequência para a qual se deseja o desenho dos primers.

Após a submissão da sequência de interesse, o programa apresentará as opções de primers possíveis de serem escolhidos.



**Primer Calculation Results**

(get more detailed information by clicking in the pairs position)

8 best Pairs (of max. 1810)							
Pair-ID	Forward Primer	Reverse Primer	Qual.	Prod. Len.	T <sub>m</sub> Diff.	FPPos.	RFPpos.
1	CGAGCTCACAGGAGCA	ACGTCGTTGCGTCGTA	883	516	0	581	1081
2	GAGCTCACAGGAGCACA	ACGTCGTTGCGTCGTA	883	515	0	582	1081
3	TCCACATAGCATGCA	TTCCACTCGCTTCTGA	883	517	5	1813	2313
4	CCAGCCCCAACTACCA	ATAGGTCGGGGTTACACA	883	518	0	2900	3400
5	CCACATAGCATGCAA	TTCCACTCGCTTCTGA	882	516	6	1814	2313
6	CCCAAAGTACGACGCAA	GTTGGGAGCTAGTACCA	882	515	1	1074	1571
7	ATCCATCACAGCAACCA	GGGGAATTGTTGTGGA	882	516	2	2623	3121
8	CAACCCCAACAACCACA	GGTTGTTGTTGTGGA	881	512	8	2757	3254

**Resultado: sequências dos primers (sensi e anti-sensi)****Pair-data of Pair-ID 1**

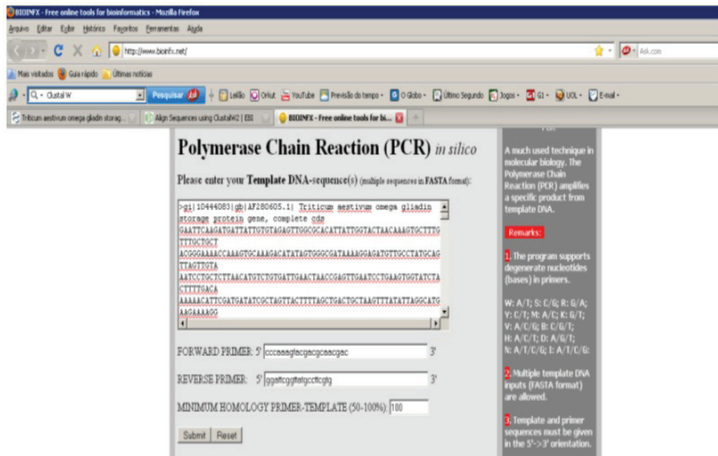
	Forward Primer Data	Reverse Primer Data
Sequence	CGAGCTCACAGGAGCA	ACGTCGTTGCGTCGTA
GC Content	63	56
Position	581	1081
Primer length [bp]	16	16
Degeneracy	0	0
3' GC	50	50
3' Degeneracy	0	0
T <sub>m</sub>	53.349	53.7432
Quality	515	515

>>> [Printable HTML-Page](#) <<<**Primer Visualisation**

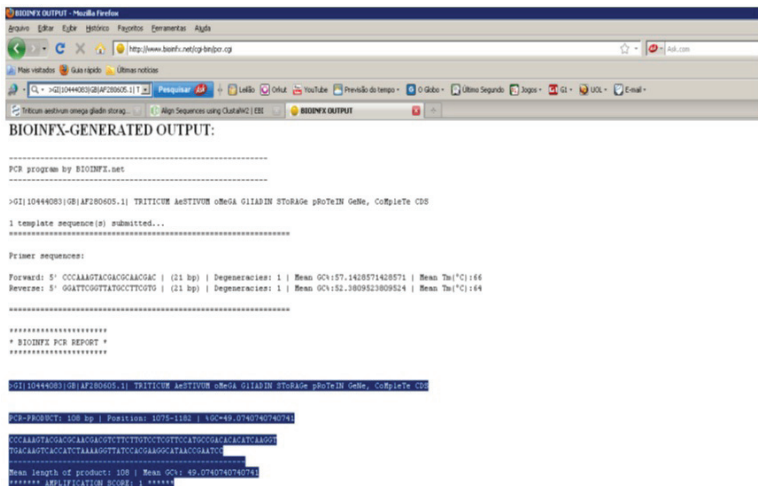
**Bioinfo.net** (<http://www.bioinfo.net/>)

É um programa de ferramentas de bioinformática livre online, que permite a simulação de uma Reação em Cadeia da Polimerase do DNA em sílica.

O processo para a obtenção do resultado ocorre basicamente de uma única maneira, a inserção da sequência de DNA e dos primers.



Posteriormente, o resultado se dará da seguinte forma:



É um programa que se faz bastante necessário por permitir uma pré-avaliação dos primers desenhados ou definidos.

## Considerações Finais

Diante do exposto, fica claro que tais ferramentas são realmente essenciais para se atingir o sucesso no desenho de iniciadores específicos, que serão usados na reação em cadeia da DNA polimerase (PCR) convencional e em tempo real.

## Referências

QU, W.; SHEN, Z.; ZHAO, D.; YANG, Y.; ZHANG, C. MFEprimer: multiple factor evaluation of the specificity of PCR primers. **Bioinformatics**, Oxford, v. 25, n. 2, p. 276-278, Jan. 2009. Disponível em: <<http://bioinformatics.oxfordjournals.org/content/early/2008/11/27/bioinformatics.btn614.full.pdf>>. Acesso em: 5 ago. 2011.

SANTOS, F. R.; ORTEGA, J. M. **Bioinformática aplicada à genômica**. Belo Horizonte, 2003. Disponível em: <<http://200.17.141.88/images/3/33/bio03.pdf>>. Acesso em: 5 jul. 2011.

## Literatura Recomendada

LI, L.-Y.; LI, Q.; YU, Y.-H.; ZHONG, M.; YANG, L.; WU, Q.-H.; QIU, Y.-R.; LUO, S.-K. A primer design strategy for PCR amplification of GC-rich DNA sequences. **Clinical Biochemistry**, v. 44, n. 8-9, p. 692-698, Jun. 2011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0009912011000634>>. Acesso em: 10 jul. 2011.

PROSDOCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A. F.; REIS, A. N. dos; JUNQUEIRA, A. C. M.; SANTOS, A. C. F. dos; NHANI JÚNIOR, A.; WUST, C. I.; CAMARGO FILHO, F.; KESSEDJIAN, J. L.; PETRETSKI, J. H.; CAMARGO, L. P.; FERREIRA, R. de G. M.; LIMA, R. P.; PEREIRA, R. M.; JARDIM, S.; SAMPAIO, V. de S.; FOLGUERAS-FLATSCHART, A. V. Bioinformática: manual do usuário. **Biotecnologia Ciência & Desenvolvimento**, Brasília, DF, ano 5, n. 29, p. 12-25, dez./jan. 2002. Disponível em: <<http://www.biotecnologia.com.br/revista/bio29/bioinf.pdf>>. Acesso em: 30 jun. 2011.

ZELTNER, D.; GLOMB, M. A.; MAEDE, D. Real-time PCR systems for the detection of the gluten-containing cereals wheat, spelt, kamut, rye, barley and oat. **European Food Research and Technology**, n. 228, n. 3, p. 321-330, Jan. 2009. Disponível em: <<http://www.springerlink.com/content/mvrn3131m4011345/fulltext.pdf>>. Acesso em: 15 jul. 2011.





---

*Agroindústria de Alimentos*

CGPE 9913



Ministério da  
Agricultura, Pecuária  
e Abastecimento

